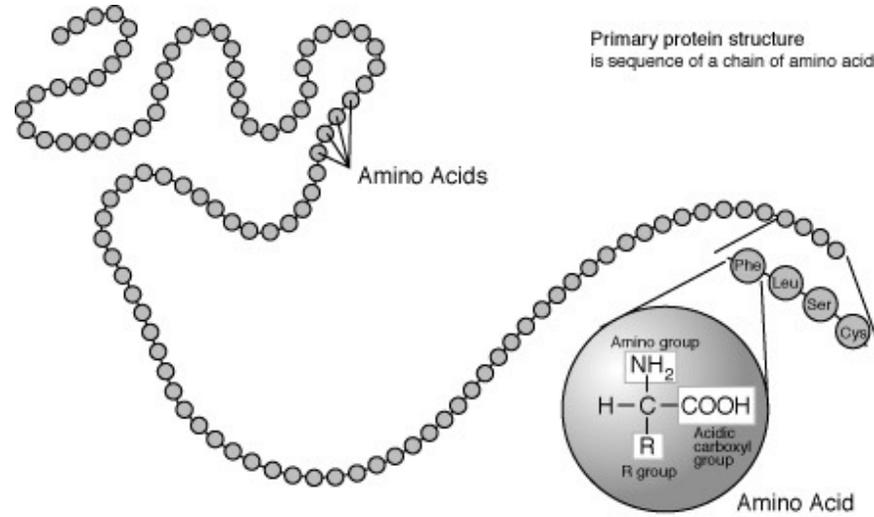
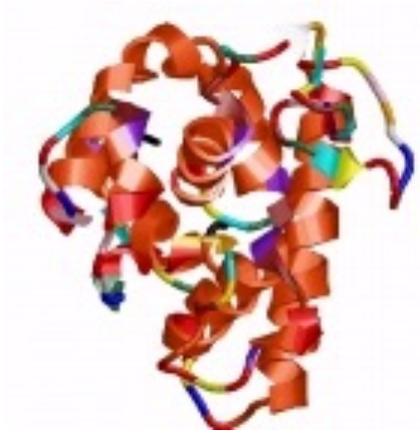


Proteomics and Mass Spectrometry

Outline

- Motivation
- Mass spectrum of a peptide
- Mass spectrometer
- Bottom-up proteomics

Protein



Primary structure is a sequence.

20 frequent amino acids.

Fold into a complex 3D structure.

A Protein Sequence

>P02769|ALBU_BOVIN Serum albumin - Bos taurus (Bovine).

```
MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIASF  
QYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASL  
RETYGDMADCCEKQEPERNECFLSHKDDSPDLPKLPDPNTLCDEFKADEKK  
FWGKYLYEIAARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLLPKIETM  
REKVLASSARQRLRCASIQKFGERALKAWSVARLSQKFPKAEFVEVTKLVD  
LTKVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKECCDKPLLEKSHC  
IAEVEKDAIPENLPPLTADFAEDKDVCKNYQEAKDAFLGSFLYEYSRRHPEY  
AVSVLLRLAKEYEATLECCAKDDPHACYSTVFDKLLKHLVDEPQNLIKQNC  
QFEKLGEGYGFQNALIVRYTRKVPQVSTPTLVEVSRS LGKVGTRCCTKPESER  
MPCTEDYLSLILNRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETY  
VPKAFDEKLFTHADICTLPDTEKQIKKQTALVELLKHKPKATEEQQLKTVME  
NFVAFVDKCCAADDKEACFAVEGPKLVVSTQTALA
```

Fundamental Questions

- **Identify, sequence, and quantify** all the proteins in a biological sample.
 - Identification: determine which proteins in a database present in the sample.
 - Sequencing: determine the amino acid sequence without needing a database.
 - Quantification: determine the quantity change of each protein under two different biological conditions.



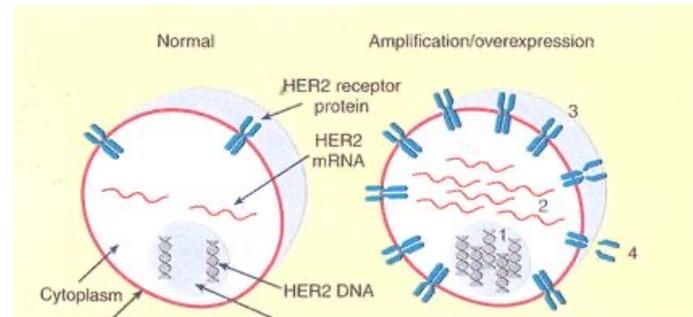
> Human Ig heavy chain variable region

```
QVQLVQSGAEVKKPGASVKVSCKASGYTFTGYYMH  
WVRQAPGQGLEWMGRINPNSGGTNYAQKFGQGRVTS  
TRDTSISTAYMELSRRLRSDDTVVYYCAR
```

Why Bother? – Example: Biomarker

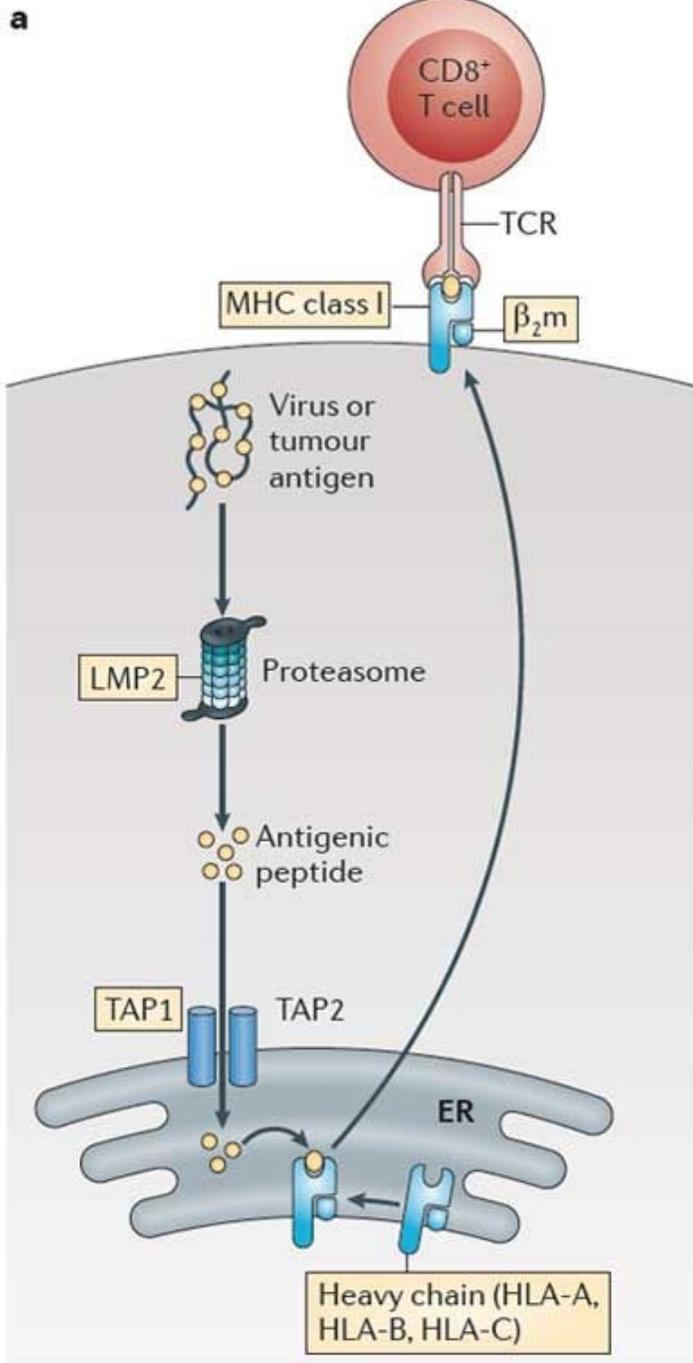
HER2-positive breast cancer is a breast cancer that tests positive for a protein called human epidermal growth factor receptor 2 (HER2), which promotes the growth of cancer cells.

Biomarker



“HER2-positive breast cancers tend to be **more aggressive** than other types of breast cancer. They're also less responsive to hormone treatment. However, treatments that **specifically target HER2** are very effective. ”



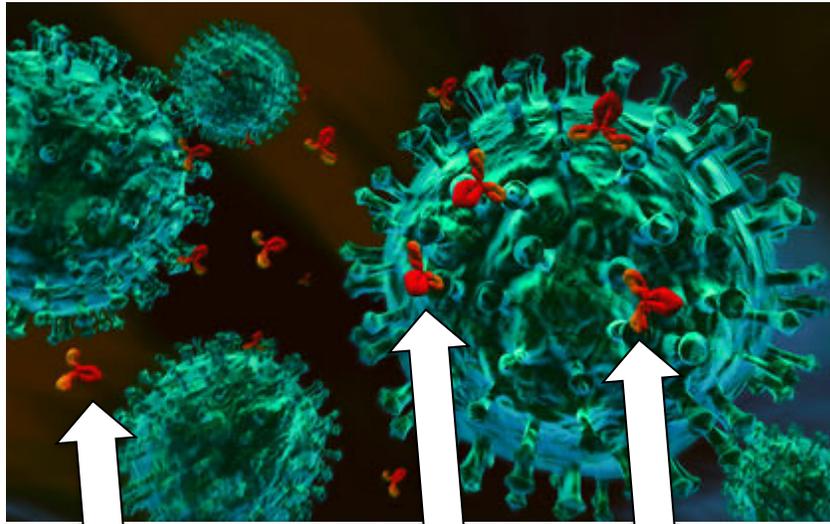


Example: Immuno peptides

- Tumor or infected cells “present” some abnormal peptides at the MHC on cell surface.
- CD8+ T cells (aka T killer cells) recognize the abnormal peptides and kill the cell.
- An actively pursued method in immunotherapy is to identify/predict the peptides presented and train the T cells to target them.
- Mass spec is the best tool to identify these peptides.

<https://nanolive.ch/immuno-oncology-t-cells-targeting-and-killing-cancer-cells/>

Example: Antibody



Antibody

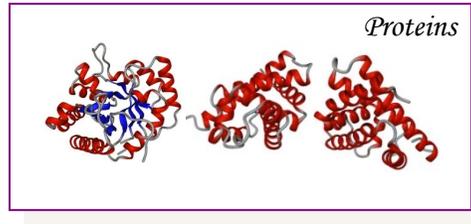
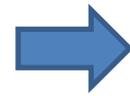


- Our immune system produces antibodies to bind to invasive pathogens (virus etc.) and cancer neoantigen.
- Antibody's amino acid sequences determine the binding target (antigen).
- Some people have stronger immune system. E.g. HIV elite controller.
- Sequencing their antibodies provides a new way to discover antibody drugs.

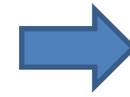
Technology Overview



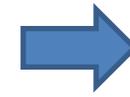
Biology sample



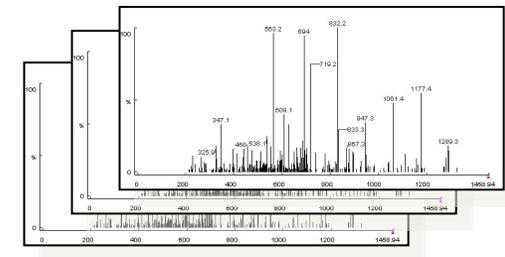
Sample Preparation



Peptides



Mass Spectrometry

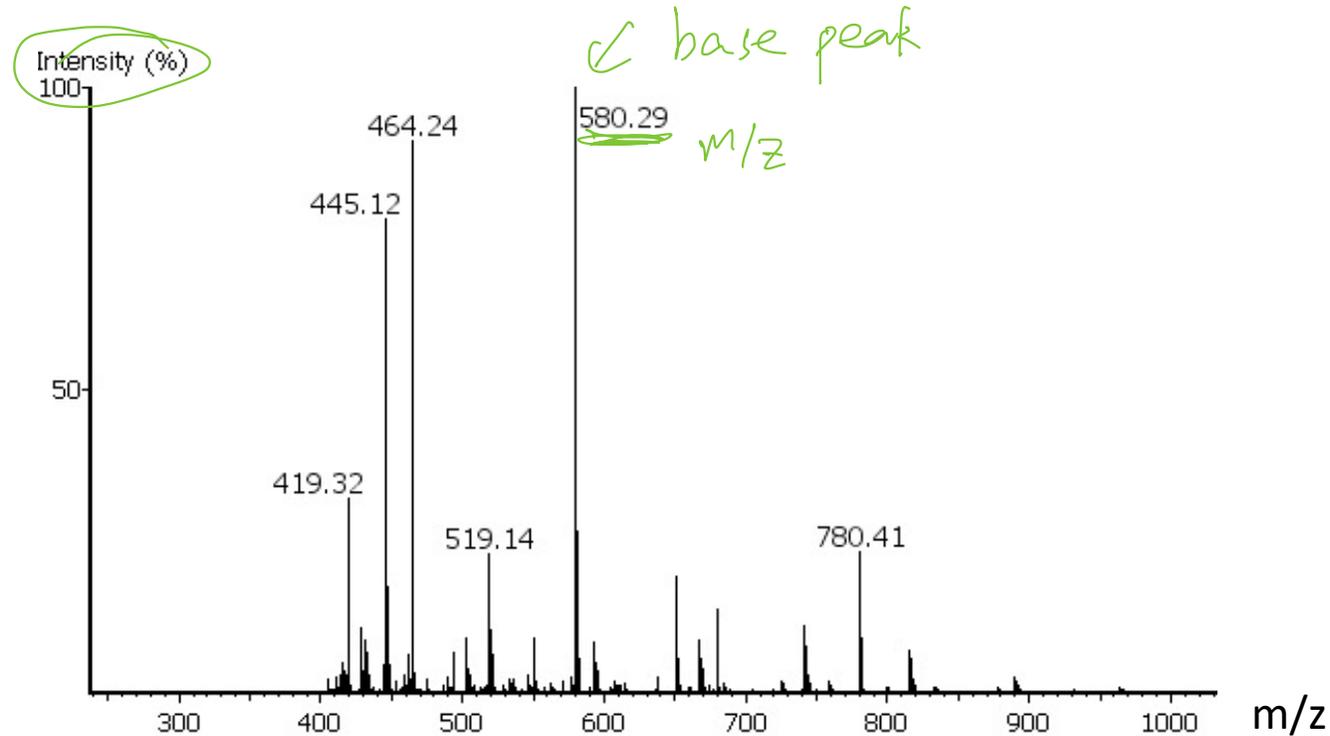


Mass Spectral Data



What're in the sample?

A Mass Spectrum



- Each peak indicates the detection of a particular type of **ion** (electrically charged molecule) with the corresponding mass to charge ratio.
- Mass of an ion can also be derived if the charge state (z) of an ion is known, which is achievable.
- A typical proteomics experiment produces tens of thousands spectra per hour.

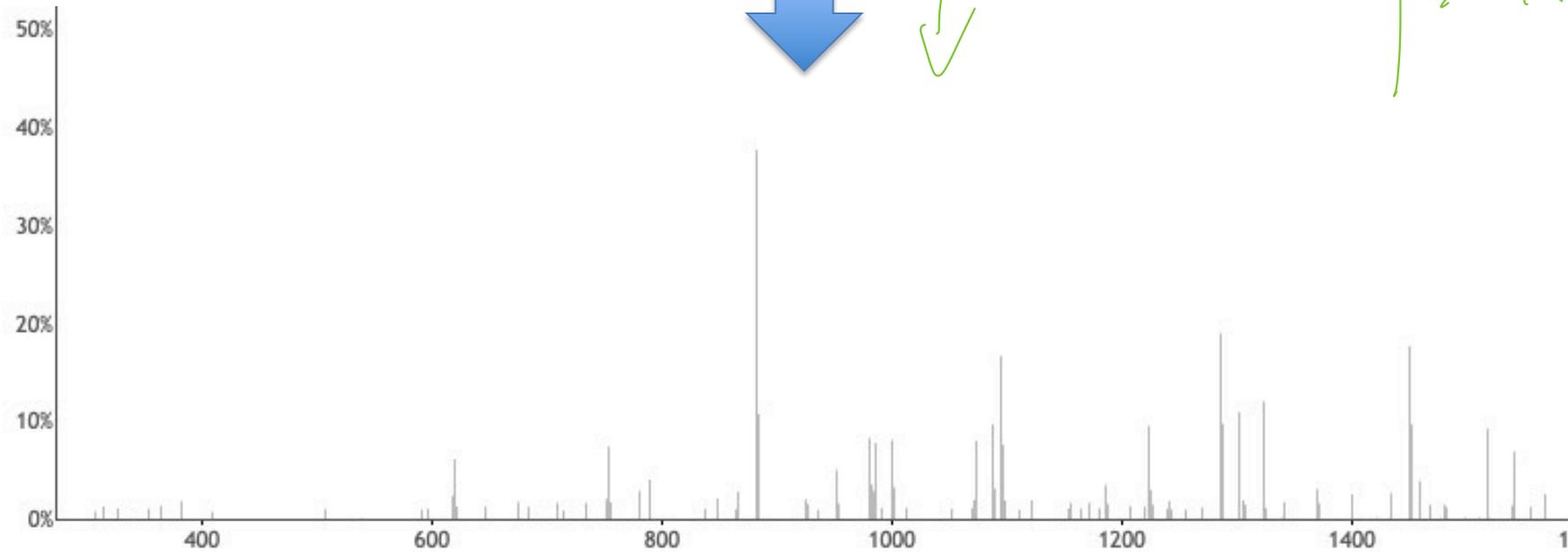
Tandem Mass (MS/MS) Spectrum of a Peptide

VVVDVSHEDPEVKF



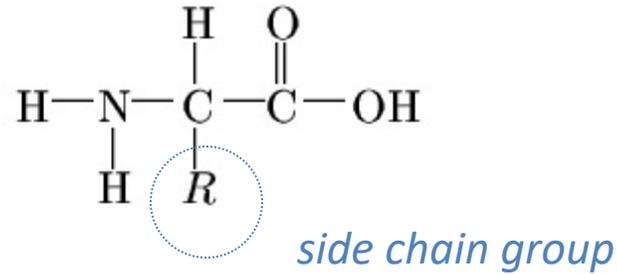
MS/MS

bioinformatics



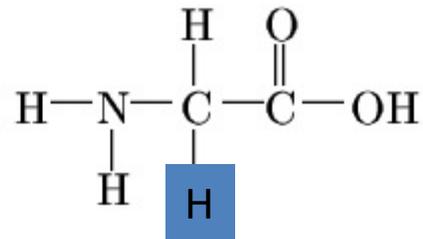
Amino Acids

- There are 20 amino acids. All have the same basic structure but with different side chains:

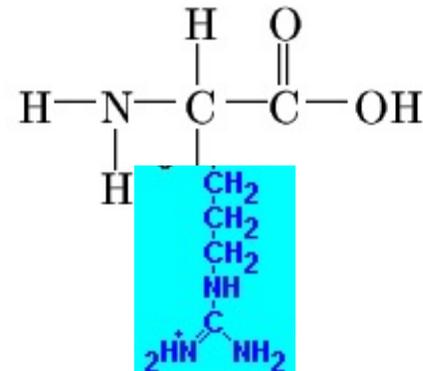


C: Carbon
H: Hydrogen
O: Oxygen
N: Nitrogen

- Examples:



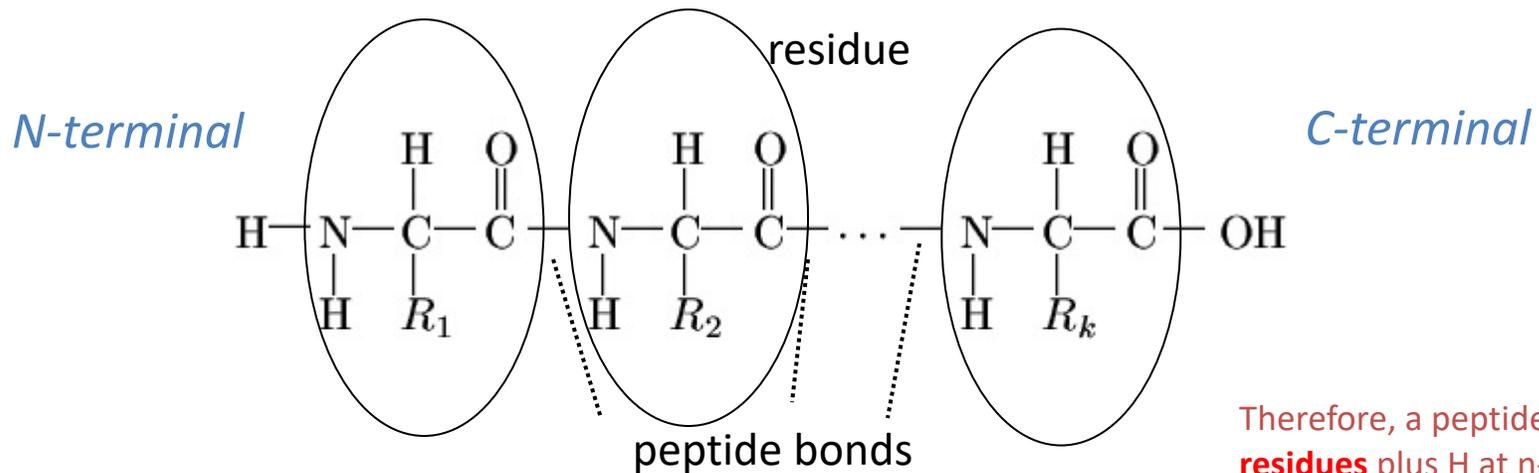
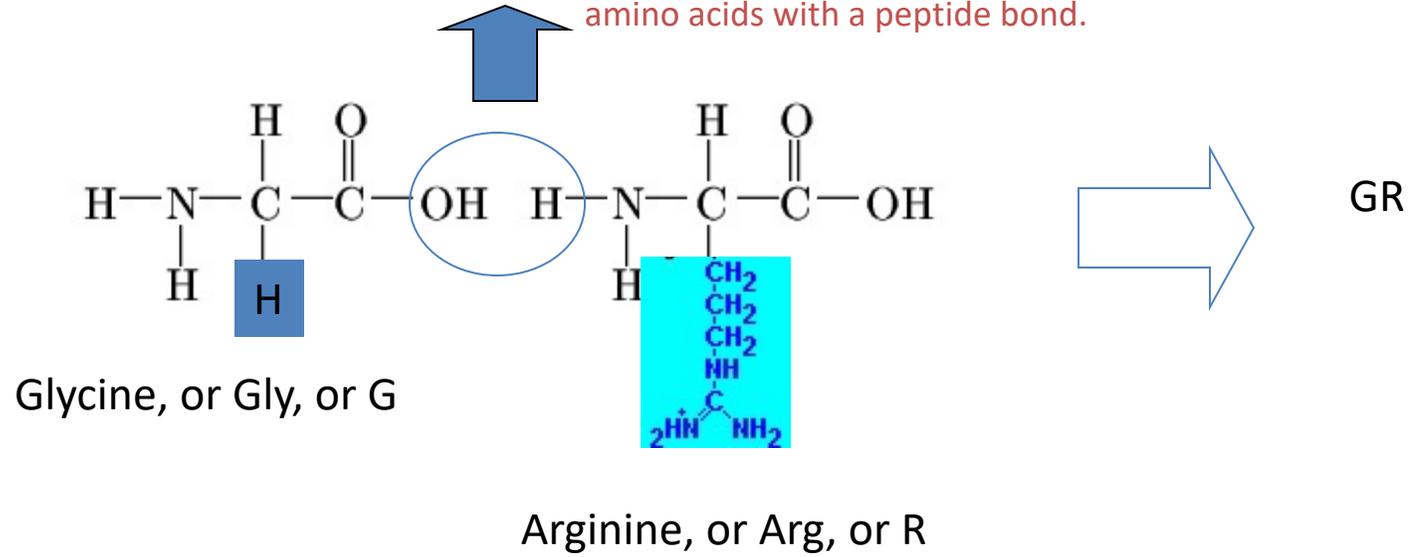
Glycine, or Gly, or G



Arginine, or Arg, or R

Peptides and Proteins

A H₂O is lost when connecting two amino acids with a peptide bond.



Therefore, a peptide is a chain of amino acid **residues** plus H at n-term and OH at c-term.

Amino Acid Residue Mass Table

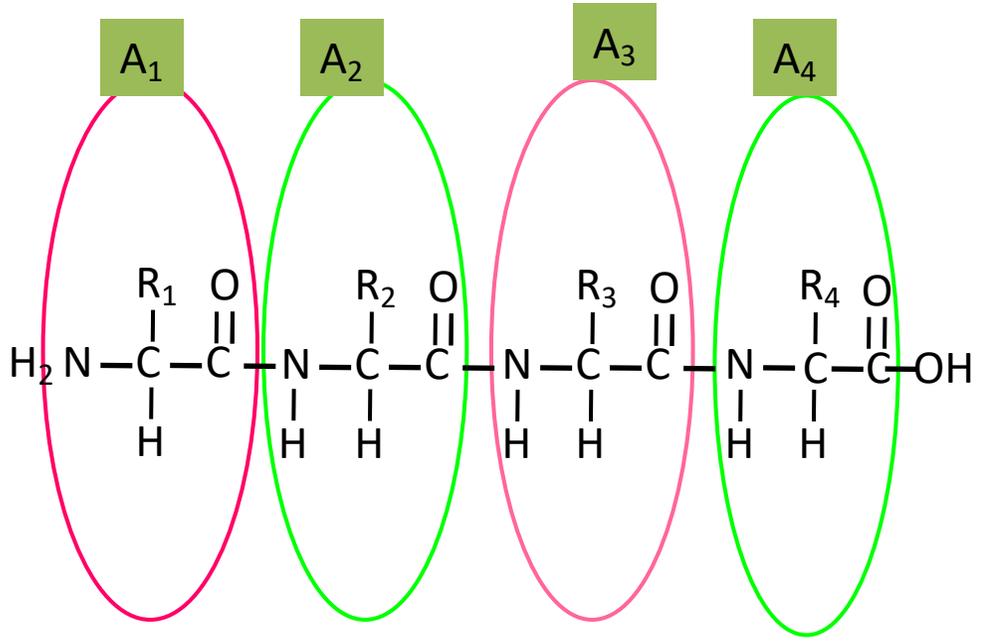
Name	3-letter code	1-letter code	Monoisotopic Mass	Average Mass	Composition
Alanine	Ala	A	71.03711	71.08	C ₃ H ₅ NO
Arginine	Arg	R	156.10111	156.2	C ₆ H ₁₂ N ₄ O
Asparagine	Asn	N	114.04293	114.1	C ₄ H ₆ N ₂ O ₂
Aspartic Acid	Asp	D	115.02694	115.1	C ₄ H ₅ NO ₃
Cysteine	Cys	C	103.00919	103.1	C ₃ H ₅ NOS
Glutamic Acid	Glu	E	129.04259	129.1	C ₅ H ₇ NO ₃
Glutamine	Gln	Q	128.05858	128.1	C ₅ H ₈ N ₂ O ₂
Glycine	Gly	G	57.02146	57.05	C ₂ H ₃ NO
Histidine	His	H	137.05891	137.1	C ₆ H ₇ N ₃ O
Isoleucine	Ile	I	113.08406	113.2	C ₆ H ₁₁ NO
Leucine	Leu	L	113.08406	113.2	C ₆ H ₁₁ NO
Lysine	Lys	K	128.09496	128.2	C ₆ H ₁₂ N ₂ O
Methionine	Met	M	131.04049	131.2	C ₅ H ₉ NOS
Phenylalanine	Phe	F	147.06841	147.2	C ₉ H ₉ NO
Proline	Pro	P	97.05276	97.12	C ₅ H ₇ NO
Serine	Ser	S	87.03203	87.08	C ₃ H ₅ NO ₂
Threonine	Thr	T	101.04768	101.1	C ₄ H ₇ NO ₂
Tryptophan	Trp	W	186.07931	186.2	C ₁₁ H ₁₀ N ₂ O
Tyrosine	Tyr	Y	163.06333	163.2	C ₉ H ₉ NO ₂
Valine	Val	V	99.06841	99.13	C ₅ H ₉ NO

Mass unit: Da (or Dalton)

$$\begin{aligned}
 &57 + 57 + 128 \\
 &= 242 \\
 &+ 18 \text{ (H}_2\text{O)} \\
 &= 260
 \end{aligned}$$

So, what's the approximate (integer) mass of peptide GGK?

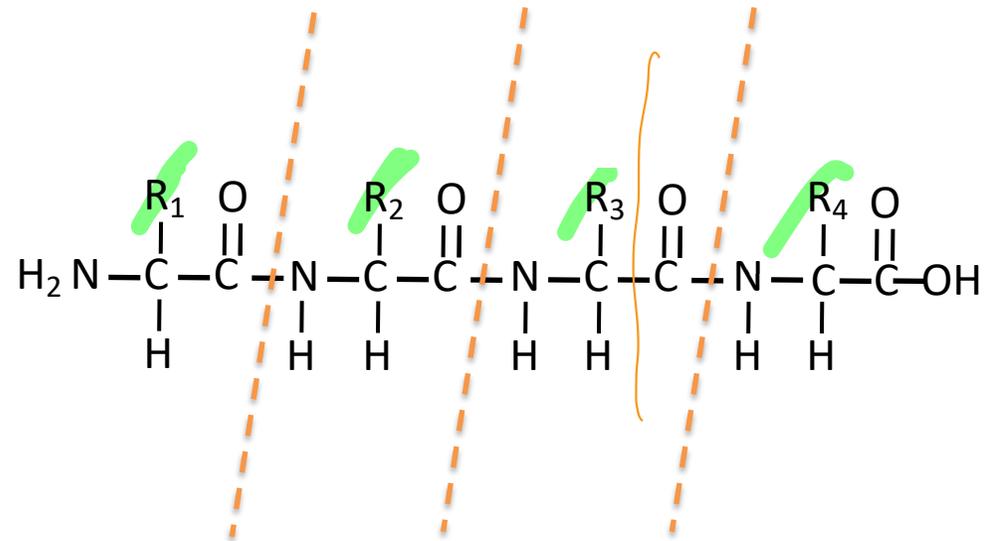
Peptide Mass



$$m(\text{A}_1)+m(\text{A}_2)+m(\text{A}_3)+m(\text{A}_4)+m(\text{H}_2\text{O})$$

$$m(\text{H}_2\text{O}) = 18.0105$$

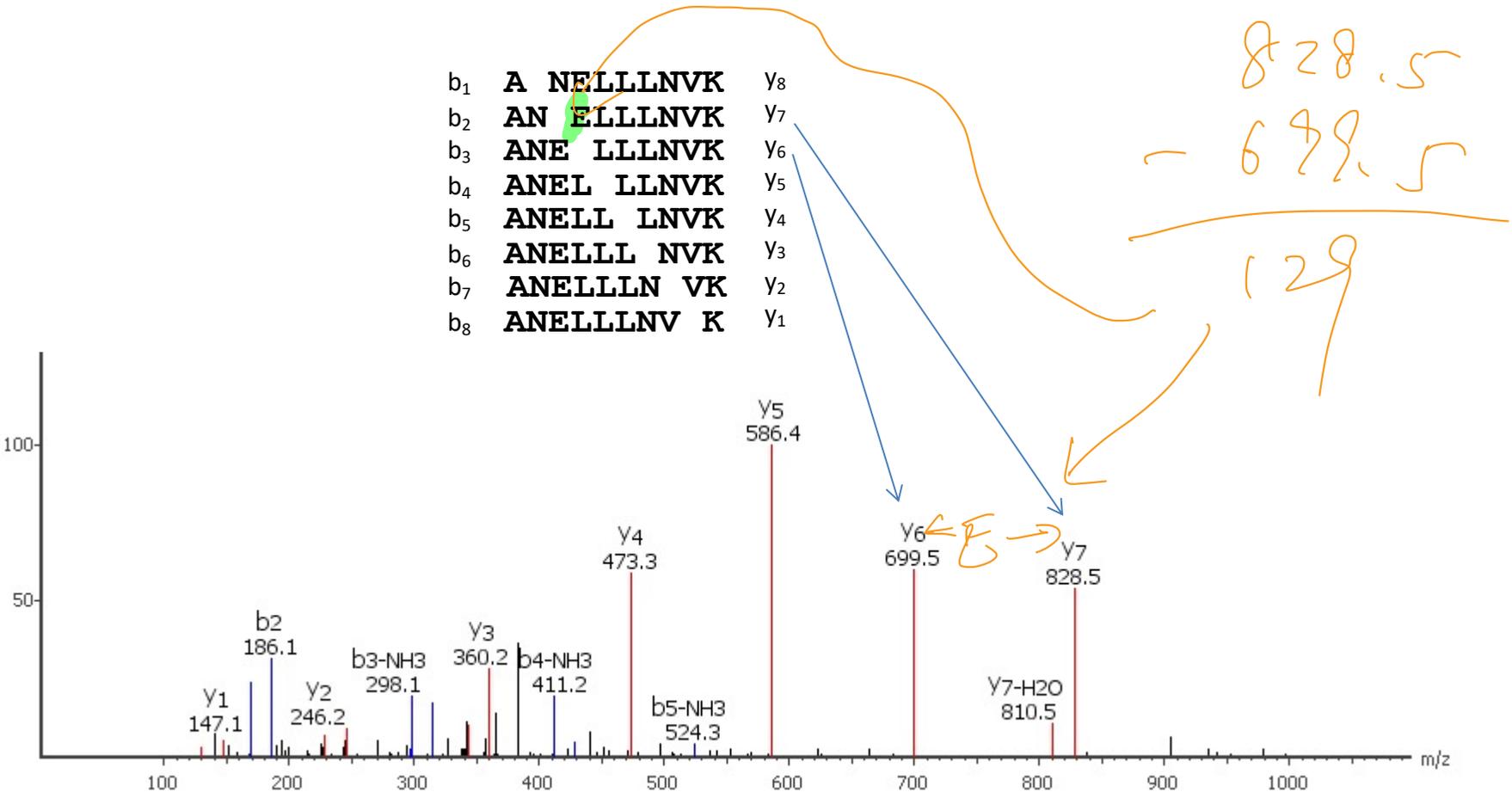
Peptide Fragmentation in MS/MS



$$m(\text{A}_1)+m(\text{A}_2)+m(\text{A}_3)+m(\text{A}_4)+m(\text{H}_2\text{O})$$

$$m(\text{H}_2\text{O}) = 18.0105$$

MS/MS of a Peptide



Mass difference between two adjacent “ladder” ions can be used to determine the amino acid (residue).